

Guaranteed Fault Detection and Isolation for Switched Affine Models

Farshad Harirchi and Sze Zheng Yong

Abstract—In the present paper, the problem of fault detection and isolation (FDI) for switched affine models is considered. We first study the model invalidation problem and its application to guaranteed fault detection. Novel and intuitive optimization-based formulations are proposed for model invalidation and T -detectability problems, which we demonstrated to be computationally more efficient than an earlier formulation that required a complicated change of variables. Moreover, we introduce a *distinguishability index* as a measure of separation between the system and fault models, which offers a practical method for finding the smallest receding time horizon that is required for fault detection, and for finding potential design recommendations for ensuring T -detectability. Then, we extend our fault detection guarantees to the problem of fault isolation with multiple fault models, i.e., the identification of the type and location of faults, by introducing the concept of I -isolability. An efficient way to implement the FDI scheme is also proposed, whose run-time does not grow with the number of fault models that are considered. Moreover, we derive bounds on detection and isolation delays and present adaptive receding horizons for reducing isolation delays. Finally, the effectiveness of the proposed method is illustrated using several examples, including an HVAC system model with multiple faults.

I. INTRODUCTION

Cyber-physical systems (CPS), i.e., systems with integrated computation, networking, and physical processes, are becoming increasingly common in our daily lives. Such systems include critical infrastructures such as traffic, power and water networks, as well as autonomous vehicles, aircrafts, home appliances and manufacturing processes. However, some major incidents involving these critical infrastructure systems as a result of cyber-attacks and system failures have taken place in the recent years and are a big source of concern. Hence, the reliability and security of CPS is paramount for their successful implementation and operation. The detection and isolation of faults and anomalies in CPS play an important role in enhancing the reliability of these systems, and in understanding the vulnerability of system components to failures and attacks.

A. Literature Review

The study of fault detection began with the introduction of the first failure detection filter by Beard in 1971 [1]. Since then, fault diagnosis has attracted a great deal of attention and has become an integral part of most, if not all system designs. Researchers have mainly approached the fault detection and isolation problem by employing either data-driven techniques or model-based approaches. These

methods can, in general, be further grouped into active and passive approaches. In active fault detection, the system is excited by a carefully designed input [2], [3], while in passive methods, the behavior of the system is not controlled or altered by the FDI scheme [4], [5].

In broad strokes, model-based fault detection and isolation schemes in the literature can be categorized into two classes, i.e., approaches that are based on residual generation and on set-membership. The former approach is more common in the fault diagnosis literature, and in this approach, the difference between the measurements and the estimates is defined as a residual or a symptom [6]. Two major trends in the residual generation techniques are the observer-based [4], [7], [8] and the parameter estimation based [9], [10] methods. Even though the residual generation based approaches are efficient and are thus widely used in the industry, their performance is highly dependent on the preciseness of the observers or the parameter estimates and also the employed residual evaluation approach. In addition, these methods do not provide any guarantees for the detection of faults. Residual-based methods are also employed for fault detection and isolation in non-linear and hybrid models [11]–[14]. In particular, an observer-based method is proposed for fault diagnosis of hybrid systems in [15], in which an extended Kalman filter is used to track the continuous behavior of the system, and a mode estimator to estimate the discrete state.

On the other hand, set-membership based fault detection and isolation techniques are proposed with the goal of providing guarantees for the detection of some specific faults. Most of these methods operate by discarding models that are not compatible with observed data, rather than identifying the most likely model. There is an extensive literature on set-membership based methods for active fault detection of linear models [16]–[18]. In [19], [20], we posed set-membership based guaranteed passive fault detection approaches for the class of switched affine models and polynomial state space models. These approaches are developed by utilizing ideas from model invalidation [21], [22] and taking advantage of recent advances in optimization. In addition, a concept called T -detectability has been introduced for finding conditions under which the fault detection scheme can be applied in a receding horizon manner without compromising detection guarantees. T -detectability is closely related to the concept of input-distinguishability of linear systems [23], [24] and mode discernibility in hybrid systems [25].

B. Main Contributions and Paper Structure

In this paper, we consider a passive fault detection and isolation scheme for switched affine systems using

This work is supported in part by DARPA grant N66001-14-1-4045. The authors are with the Electrical Engineering and Computer Science Department, University of Michigan, Ann Arbor, MI, 48109. (email: {harirchi, szyong}@umich.edu)

an optimization-based model invalidation framework, that improves and expands the results of [19] on fault detection. We provide novel formulations of the model invalidation and T -detectability algorithms that we demonstrated to be noticeably faster than the previous formulation in [19] and that have the added advantage of simplicity as no complicated change of variables are needed. Furthermore, we introduce a measure of separation between models, called *distinguishability index*. By reformulating the T -detectability optimization problem in [19], we compute the distinguishability index as a byproduct, when solving this problem. This index offers a practical way to find out if a finite receding time horizon exists, and suggests potential design options for ensuring T -detectability.

We then consider the fault isolation problem using model invalidation and introduce the concept of I -isolability when multiple faults are present. Similar to fault detection, we propose a computationally efficient optimization problem to check whether a given set of fault models is I -isolable or not. Moreover, we propose a fault diagnosis scheme that not only detects the occurrence of a fault, but also outputs a list of potential faults along with their associated ‘likelihoods’ in the form of distinguishability indices. Further, a theoretical analysis for bounds on detection and isolation delays is provided and an adaptive fault isolation scheme is proposed to reduce isolation delays. The run-time of our FDI scheme also does not grow with the number of fault models. Finally, these results are illustrated using a numerical model of a heating, ventilating, and air conditioning (HVAC) system.

II. PRELIMINARIES

In this section, the notation used throughout the paper and the modeling framework we consider are described.

A. Notation

Let $\mathbf{x} \in \mathbb{R}^n$ denote a vector and $\mathbf{M} \in \mathbb{R}^{n \times m}$ represent a matrix. The infinity norm of a vector \mathbf{x} is denoted by $\|\mathbf{x}\| = \max_i |\mathbf{x}^i|$, where \mathbf{x}^i denotes the i^{th} element of vector \mathbf{x} . The set of positive integers up to n is denoted by \mathbb{Z}_n^+ , and the set of non-negative integers up to n is denoted by \mathbb{Z}_n^0 .

B. Modeling Framework

In this paper, we consider systems that can be represented by discrete-time switched affine (SWA) models.

Definition 1: (SWA Model) A switched affine model is defined by:

$$\mathcal{G} = (\mathcal{X}, \mathcal{E}, \mathcal{U}, \{G_i\}_{i=1}^m), \quad (1)$$

where $\mathcal{X} \subset \mathbb{R}^n$ is the set of states, $\mathcal{E} \subset \mathbb{R}^{n_y + n_p}$ is the set of measurement and process noise signals, $\mathcal{U} \subset \mathbb{R}^{n_u}$ is the set of inputs and $\{G_i\}_{i=1}^m$ is a collection of m modes. For all $i \in \mathbb{Z}_m^+$, the i^{th} mode is an affine model:

$$G_i = \{\mathbf{A}_i, \mathbf{B}_i, \mathbf{C}_i, \mathbf{D}_i, \mathbf{f}_i, \mathbf{g}_i\}. \quad (2)$$

The evolution of \mathcal{G} is governed by:

$$\begin{aligned} \mathbf{x}_{t+1} &= \mathbf{A}_{\sigma_t} \mathbf{x}_t + \mathbf{B}_{\sigma_t} \mathbf{u}_t + \mathbf{f}_{\sigma_t} + \boldsymbol{\nu}_t, \\ \mathbf{y}_t &= \mathbf{C}_{\sigma_t} \mathbf{x}_t + \mathbf{D}_{\sigma_t} \mathbf{u}_t + \mathbf{g}_{\sigma_t} + \boldsymbol{\eta}_t, \end{aligned} \quad (3)$$

where $\boldsymbol{\nu} \in \mathbb{R}^{n_p}$ and $\boldsymbol{\eta} \in \mathbb{R}^{n_y}$ denote the process and measurement noise signals, respectively, and σ_t indicates the active mode at time t .

Remark 1: We assume $\mathcal{X}, \mathcal{E}, \mathcal{U}$ are convex and compact sets. In particular, we consider the following form for the admissible sets:

$$\begin{aligned} \mathcal{X} &= \{\mathbf{x} \mid P\mathbf{x} \leq p\}, \quad \mathcal{E} = \{[\boldsymbol{\eta}^T \ \boldsymbol{\nu}^T]^T \mid \|\boldsymbol{\eta}\| \leq \epsilon_\eta, \|\boldsymbol{\nu}\| \leq \epsilon_\nu\} \\ \mathcal{U} &= \{\mathbf{u} \mid \|\mathbf{u}\| \leq U\}, \end{aligned} \quad (4)$$

where $P \in \mathbb{R}^{n_p \times n}$ and $p \in \mathbb{R}^{n_p}$. Note that our analysis holds true for any $\mathcal{X}, \mathcal{E}, \mathcal{U}$ that are convex sets, but for the sake of simplicity in notation, we use the above mentioned admissible sets.

We define the fault model as follows:

Definition 2 (Fault Model): A fault model for a switched affine system $\mathcal{G} = (\mathcal{X}, \mathcal{E}, \mathcal{U}, \{G_i\}_{i=1}^m)$ is another switched affine model $\bar{\mathcal{G}} = (\bar{\mathcal{X}}, \bar{\mathcal{E}}, \bar{\mathcal{U}}, \{\bar{G}_i\}_{i=1}^{\bar{m}})$ with the same number of states, inputs and outputs.

Further, to describe our framework of model invalidation and T -detectability for fault detection and isolation in the next section, we define the following.

Definition 3 (Length- N behavior): The length- N behavior associated with an SWA system \mathcal{G} is the set of all length- N input-output trajectories compatible with \mathcal{G} , given by the following set:

$$\begin{aligned} \mathcal{B}_{swa}^N(\mathcal{G}) &:= \{ \{ \mathbf{u}_t, \mathbf{y}_t \}_{t=0}^{N-1} \mid \mathbf{u}_t \in \mathcal{U} \text{ and } \exists \mathbf{x}_t \in \mathcal{X}, \sigma_t \in \mathbb{Z}_m^+, \\ &\quad [\boldsymbol{\eta}_t^T \ \boldsymbol{\nu}_t^T]^T \in \mathcal{E}, \text{ for } t = 0, \dots, N-1 \text{ s.t. (3) holds} \}. \end{aligned}$$

Moreover, with a slight abuse of terminology, we will call $\mathcal{B}_{swa}^N(\mathcal{G})$ the *behavior* of the system \mathcal{G} for conciseness.

III. MODEL INVALIDATION AND T -DETECTABILITY

A. Model Invalidation

In our previous work [19], [20], we established a theoretical framework that can be utilized in order to develop fault detection schemes based on the achievements in model invalidation, a framework that we will also consider in this paper. The model invalidation problem is to check whether some given data can be represented by a model or not. More formally, the model invalidation problem is as follows:

Problem 1 (Model Invalidation): Given an SWA model \mathcal{G} and an input-output sequence $\{ \mathbf{u}_t, \mathbf{y}_t \}_{t=0}^{N-1}$, determine whether or not the input-output sequence is contained in the behavior of \mathcal{G} , i.e., whether or not the following is true:

$$\{ \mathbf{u}_t, \mathbf{y}_t \}_{t=0}^{N-1} \in \mathcal{B}_{swa}^N(\mathcal{G}). \quad (5)$$

Clearly, if the model is invalidated by data, i.e., (5) does not hold, and the model is precise, it is equivalent to the data representing an abnormal behavior. Hence, model invalidation can be used as a fault detection scheme. Moreover, our previous work [19] has shown that model invalidation problem for SWA models can be posed as a Mixed-Integer Linear Program (MILP) feasibility check problem.

In this paper, we propose a new MILP formulation, that we believe is much more intuitive and computationally efficient. We obtain this novel formulation by taking advantage of Special Ordered Set of degree 1 (SOS-1) constraints [26],

that are readily implementable in most off-the-shelf optimization softwares. In brief, an SOS-1 constraint is a set of variables for which at most one variable in the set may be non-zero. Our new formulation is cleaner because this type of constraints allows us to formulate the feasibility check problem without introducing complicated change of variables as was previously done in [19]. Moreover, SOS-1 constraints, which are by nature integral constraints, make the branch and bound search procedures noticeably faster (see, e.g., [27, Section 3.3.4] for a discussion). Our new model invalidation problem using SOS-1 constraints is presented below, which we will demonstrate to be much faster than an earlier formulation [19], [28] in Section VI-A.

Proposition 1: Given an SWA model \mathcal{G} and an input-output sequence $\{\mathbf{u}_t, \mathbf{y}_t\}_{t=0}^{N-1}$, the model is invalidated if and only if the following problem is infeasible.

$$\begin{aligned} & \text{Find } \mathbf{x}_t, \boldsymbol{\eta}_t, \boldsymbol{\nu}_t, a_{i,t}, \mathbf{s}_{i,t}, \mathbf{r}_{i,t} \text{ for } \forall t \in \mathbb{Z}_{N-1}^0, \forall i \in \mathbb{Z}_m^+ \\ & \text{s.t. } \forall j \in \mathbb{Z}_n^+, \forall k \in \mathbb{Z}_{n_y}^+, \forall l \in \mathbb{Z}_{n_p}^+, \forall t \in \mathbb{Z}_{N-1}^0, \text{ we have:} \\ & \quad \mathbf{x}_{t+1} = \mathbf{A}_i \mathbf{x}_t + \mathbf{B}_i \mathbf{u}_t + \mathbf{f}_i + \boldsymbol{\nu}_t + \mathbf{s}_{i,t} \\ & \quad \mathbf{y}_t = \mathbf{C}_i \mathbf{x}_t + \mathbf{D}_i \mathbf{u}_t + \mathbf{g}_i + \boldsymbol{\eta}_t + \mathbf{r}_{i,t} \quad (\text{P}_{MI}) \\ & \quad P \mathbf{x}_t \leq p, a_{i,t} \in \{0, 1\}, \sum_{i \in \mathbb{Z}_m^+} a_{i,t} = 1, \|\boldsymbol{\nu}_t\| \leq \epsilon_\nu, \\ & \quad \|\boldsymbol{\eta}_t\| \leq \epsilon_\eta, (a_{i,t}, \mathbf{s}_{i,t}^j) : \text{SOS-1}, (a_{i,t}, \mathbf{r}_{i,t}^k) : \text{SOS-1} \end{aligned}$$

where $\mathbf{s}_{i,t}$ and $\mathbf{r}_{i,t}$ are slack variables that are free when $a_{i,t}$ is zero and zero otherwise. We refer to this problem as $\text{Feas}(\{\mathbf{u}_t, \mathbf{y}_t\}_{t=0}^{N-1}, \mathcal{G})$.

Intuitively, the infeasibility of (P_{MI}) indicates that there are no state, input and noise values that can generate input-output sequence from the model, and hence it is impossible that the data is generated by the model. Proposition 1 enables us to solve the model invalidation problem by checking the feasibility of (P_{MI}) , which is a MILP with SOS-1 constraints that can be efficiently solved with many off-the-shelf softwares, e.g., [29], [30].

B. T -Detectability

The model invalidation problem can be solved for the input-output sequence of any given time horizon to detect faults, but the number of variables and constraints increase with the size of the time horizon. Thus, a few questions naturally arise with regards to this time horizon.

First, one may ask if the smallest receding time horizon T can be found, for which two different models are guaranteed to be distinguishable. This question leads us to define the notion of T -detectability, previously presented in [19]. T -detectability is defined for a pair of system and fault models, which means that the trajectory generated from the two models cannot be identical for a time horizon of length T for any initial state and any noise signals. This notion is very similar to the concept of input-distinguishability, which is defined for linear time-invariant models in [23], [24]. T -detectability is formally defined as follows:

Definition 4 (T -detectability): A fault model $\bar{\mathcal{G}}$ for a switched affine system \mathcal{G} is called T -detectable if $\mathcal{B}_{swa}^T(\mathcal{G}) \cap \mathcal{B}_{swa}^T(\bar{\mathcal{G}}) = \emptyset$, where T is a positive integer.

Thus, given two SWA models and an integer T , the T -detectability problem is to check whether the two models are

T -detectable or not. This problem can be addressed using a Satisfiability Modulo Theory approach [19], or a MILP feasibility check [28]. As with model invalidation in the previous section, we will also propose an alternative MILP formulation for checking T -detectability, which employs SOS-1 type constraints and as before, is more intuitive and computationally superior (cf. Section VI-A). Note that in the following T -detectability test, we have added a decision variable δ that will be important in a later discussion, which can be computed with little additional computational cost.

Theorem 1: The fault model $\bar{\mathcal{G}}$ is T -detectable for the system model \mathcal{G} if and only if the following problem is infeasible.

$$\begin{aligned} & \bar{\delta} = \min_{\mathbf{x}, \bar{\mathbf{x}}, \mathbf{u}, \boldsymbol{\eta}, \boldsymbol{\nu}, \bar{\boldsymbol{\nu}}, \mathbf{s}, \bar{\mathbf{s}}, \mathbf{r}, \bar{\mathbf{r}}, a, \bar{a}, \delta} \delta \\ & \text{s.t. } \forall t \in \mathbb{Z}_{T-1}^0, \forall i \in \mathbb{Z}_m^+, \forall j \in \mathbb{Z}_m^+, \forall k \in \mathbb{Z}_n^+, \forall l \in \mathbb{Z}_{n_y}^+ \\ & \quad \forall h \in \mathbb{Z}_{n_p}^+, \bar{h} \in \mathbb{Z}_{n_{\bar{p}}}^+ \\ & \quad \mathbf{x}_{t+1} = \mathbf{A}_i \mathbf{x}_t + \mathbf{B}_i \mathbf{u}_t + \mathbf{f}_i + \boldsymbol{\nu}_t + \mathbf{s}_{i,t} \\ & \quad \bar{\mathbf{x}}_{t+1} = \bar{\mathbf{A}}_j \bar{\mathbf{x}}_t + \bar{\mathbf{B}}_j \mathbf{u}_t + \bar{\mathbf{f}}_j + \bar{\boldsymbol{\nu}}_t + \bar{\mathbf{s}}_{j,t} \\ & \quad P \mathbf{x}_t \leq p, \bar{P} \bar{\mathbf{x}}_t \leq \bar{p}, \quad (\text{P}_T) \\ & \quad \mathbf{C}_i \mathbf{x}_t + \mathbf{D}_i \mathbf{u}_t + \mathbf{g}_i + \boldsymbol{\eta}_t = \bar{\mathbf{C}}_j \bar{\mathbf{x}}_t + \bar{\mathbf{D}}_j \mathbf{u}_t + \bar{\mathbf{g}}_j + \bar{\boldsymbol{\eta}}_t + \mathbf{r}_{i,t}, \\ & \quad a_{i,j,t} \in \{0, 1\}, \sum_{i \in \mathbb{Z}_m^+} \sum_{j \in \mathbb{Z}_m^+} a_{i,j,t} = 1 \\ & \quad \|\boldsymbol{\eta}_t\| \leq \epsilon_\eta, \|\bar{\boldsymbol{\eta}}_t\| \leq \epsilon_{\bar{\eta}}, \|\boldsymbol{\nu}_t\| \leq \epsilon_\nu, \|\bar{\boldsymbol{\nu}}_t\| \leq \epsilon_{\bar{\nu}}, \|\mathbf{u}_t\| \leq U, \\ & \quad (a_{i,j,t}, \mathbf{s}_{i,t}^k) : \text{SOS-1}, (a_{i,j,t}, \bar{\mathbf{s}}_{j,t}^k) : \text{SOS-1} \\ & \quad (a_{i,j,t}, \mathbf{r}_{i,j,t}^l) : \text{SOS-1}, \left\| \begin{bmatrix} \boldsymbol{\eta}_t \\ \boldsymbol{\nu}_t \end{bmatrix} - \begin{bmatrix} \bar{\boldsymbol{\eta}}_t \\ \bar{\boldsymbol{\nu}}_t \end{bmatrix} \right\| \leq \delta. \end{aligned}$$

We refer to the above-mentioned problem as $\text{Feas}_T(\mathcal{G}, \bar{\mathcal{G}})$.

Proof: Except for the last constraint, this is an equivalent formulation to the MILP feasibility problem of Theorem 1 in [28]. Clearly, the last constraint does not change the feasible set, therefore the feasibility of (P_T) is necessary and sufficient for T -detectability. ■

While Theorem 1 enables us to solve the T -detectability problem, if the two models are not T -detectable, i.e., the solution to P_T is feasible, it additionally delivers $\bar{\delta}$, which we argue is a good indication and measure for the separability of two models. In essence, $\bar{\delta}$ can be interpreted as the noise effort that is required to make the trajectories of the two models identical. A larger value for $\bar{\delta}$ indicates a larger separation between the two models that the noise has to compensate for. Hence, we will refer to the normalized version of $\bar{\delta}$ as the *distinguishability index*, given by

$$\delta^* = \frac{\bar{\delta}}{\delta_{\max}}, \quad (6)$$

where $\delta_{\max} \doteq \min\{\max\{\epsilon_\eta + \epsilon_{\bar{\eta}}, \epsilon_\nu + \epsilon_{\bar{\nu}}\}, \max\{\epsilon_\eta, \epsilon_{\bar{\eta}}\} + \max\{\epsilon_\nu, \epsilon_{\bar{\nu}}\}\}$ is an upper bound on $\bar{\delta}$; hence, $0 \leq \delta^* \leq 1$.

Moreover, to find the smallest T for which we have T -detectability, one could iterate with T increasing from 1 until the T -detectability problem in Theorem 1 becomes infeasible. But, if δ^* is small for some T , then it may make sense to consider larger increases in T to speed up computation. Thus, δ^* can be used as a heuristic for choosing the next T to solve the T -detectability problem. In addition, one may also ask about when the iterations with increasing T can be terminated with some confidence that a finite T

does not exist. Once again, we can consider the trend of δ^* with increasing T and terminate the iterations when δ^* reaches a plateau. This will be demonstrated to be effective in a simulation example in Section VI-C. In addition, when this index reaches a (non-zero) plateau and the problem remains not T -detectable, then it would be possible to use any value that is smaller than the maximum δ^* to derive the maximum allowed uncertainty for a system such that fault detection is guaranteed. This may suggest possible design remedies involving the choice of sensors with better precision or the employment of noise isolation platforms to reduce the amount of noise, in order to facilitate fault detection.

IV. FAULT ISOLATION

A great deal of attention has been devoted to fault isolation, because of its important role in determining the source of faults, which can in turn save a significant amount of effort in accommodating the detected faults. Here, we propose an approach to the fault isolation problem for switched affine models, which provides necessary and sufficient conditions for fault isolation. In addition, this approach can be implemented relatively fast, thanks to the recent advances in the mixed-integer linear programming tools [30].

A. I -Isolability

In this section, we propose mixed integer linear certificates to ensure the isolability of two faults, formally defined below.

Definition 5 (I -isolability): Two fault models $\bar{\mathcal{G}}_1$ and $\bar{\mathcal{G}}_2$ are called I -step isolable or simply I -isolable, if $\mathcal{B}_{swa}^I(\bar{\mathcal{G}}_1) \cap \mathcal{B}_{swa}^I(\bar{\mathcal{G}}_2) = \emptyset$, where I is a positive integer.

With this definition, it is clear that I -isolability and T -detectability are identical, if we replace the system model in the T -detectability problem with one of the fault models. Hence, we can use Theorem 1 with a slight modification to check the I -isolability for two fault models.

Proposition 2: Given two fault models $\bar{\mathcal{G}}_1$ and $\bar{\mathcal{G}}_2$, the two faults are I -isolable if and only if $\text{Feas}_I(\bar{\mathcal{G}}_1, \bar{\mathcal{G}}_2)$ (as defined in Theorem 1 with I , $\bar{\mathcal{G}}_1$ and $\bar{\mathcal{G}}_2$ in place of T , \mathcal{G} and $\bar{\mathcal{G}}$) is infeasible.

Proposition 2 enables us to check whether two faults are I -isolable or not in a tractable way. Once, we have confirmed that two faults are I -isolable, it is guaranteed that by observing only I samples of the data, we can isolate them.

B. Calculating the smallest I

Similar to what we proposed for calculating the smallest T -detectability index in [28], we can calculate the smallest isolability index. That is by starting from $I = 1$ and increasing the time horizon until the two models are I -isolable. However, as discussed in Section III-B, we may similarly use the *distinguishability index* δ^* as a heuristic for speeding up this computation.

C. Multiple-Fault Scenario

The fault isolation problem becomes marginally more challenging in the case of multiple fault models. Let us assume that there exist N_f fault models for a specific system. We consider the following assumptions:

A1 (*Detectability Assumption*): We assume that $\forall j \in \mathbb{Z}_{N_f}^+$, there exists a finite T_j such that $\bar{\mathcal{G}}_j$ is T_j -detectable for the system \mathcal{G} .

A2 (*Isolability Assumption*): We assume that $\forall m, n \in \mathbb{Z}_{N_f}^+$, $m \neq n$, there exists a finite $I_{m,n}$ such that $\bar{\mathcal{G}}_m$ and $\bar{\mathcal{G}}_n$ are $I_{m,n}$ -isolable.

Remark 2: The detectability and isolability assumptions above are definitely strong assumptions, however, they are necessary and sufficient conditions for providing detection and isolation guarantees for passive fault detection approaches with a receding horizon. In fact, these assumptions are also typical assumptions in passive fault detection, and this was the impetus for considering active fault diagnosis methods, e.g., [31], [32], which, at the cost of perturbing the desired input to the system, make fault diagnosis possible for a wider class of faults. Moreover, as we will show with examples in Section VI, these assumptions hold for many parametric fault scenarios in real-world applications.

Now, we will define I -isolability for multiple faults.

Definition 6 (I -isolability for multiple faults): Consider a set of N_f fault models. This set is called I -isolable if all pairs of fault models, m, n in the set are $I_{m,n}$ -isolable for some finite $I_{m,n}$.

Proposition 3: Consider N_f fault models. Under assumption 2, the set of faults is I -isolable, if and only if

$$I = \max_{m, n, m \neq n} I_{m, n}.$$

is a finite integer.

Proof: Suppose that the set of faults is I -isolable, then any two faults can be isolated in I steps. Hence, $I_{m,n}$ for all possible pairs of faults are finite and at most I . On the other hand, if all fault pairs are $I_{m,n}$ -isolable, then there exists a maximum for all isolability indices, I , which guarantees the isolability of the set of faults. ■

V. FDI SCHEME

In this section, we propose an FDI scheme, which consists of two steps:

- 1) *Off-line step:* In the off-line step, under Assumptions A1 and A2, we calculate the following quantities:

$$\text{Isolability index: } I = \max_{m, n} I_{m, n}, \quad m, n \in \mathbb{Z}_{N_f}^+, \quad m \neq n,$$

$$\text{Isolability index for fault } i: \tilde{I}_i = \max_{j \in \mathbb{Z}_{N_f}^+, j \neq i} I_{i, j},$$

$$\text{Detectability index: } T = \max_{j \in \mathbb{Z}_{N_f}^+} T_j.$$

$$\text{Length of memory: } K = \max\{T, I\}$$

- 2) *On-line step:* In this step, we leverage $N_f + 1$ parallel monitors corresponding to system and fault models. The monitors are labeled as $\{\mathcal{M}_0, \mathcal{M}_1, \dots, \mathcal{M}_{N_f}\}$, where \mathcal{M}_0 corresponds to the system model and \mathcal{M}_i corresponds to the i^{th} fault model. First, only \mathcal{M}_0 is active for fault detection. The rest of the monitors will be “off” until a fault is detected by \mathcal{M}_0 . The inputs to each monitor at time t are the input-output sequence of length $K_i = \max\{\tilde{I}_i, T_i\}$, $\{\mathbf{u}_k, \mathbf{y}_k\}_{k=t-K_i+1}^t$, and the corresponding model $\bar{\mathcal{G}}_i$. For instance, \mathcal{M}_0 knows \mathcal{G} ,

and at each time step, it solves the model invalidation problem, $Feas(\{\mathbf{u}_k, \mathbf{y}_k\}_{k=t-T+1}^t, \mathcal{G})$. If the problem is feasible, the monitor outputs 0, otherwise it outputs 1. In the latter case, the bank of fault monitors is activated and *parallelly* solves the model invalidation problems for all fault models, i.e., to check if \mathcal{M}_j solves $Feas(\{\mathbf{u}_k, \mathbf{y}_k\}_{k=t-K_j+1}^t, \bar{\mathcal{G}}_j)$ for each $j \in \mathbb{Z}_{N_f}^+$. By Assumptions A1 and A2, it is guaranteed that in this case, the problem of at most one monitor is feasible. The output block receives the signal from all the monitors and shows two elements, the first element is 1, which indicates that a fault has occurred, and the second element is $k_f \in \mathbb{Z}_{N_f}^+$ if the fault matches k_f^{th} fault model, or 0 if the fault does not match any of the fault models.

Such an FDI scheme is illustrated in Fig. 1.

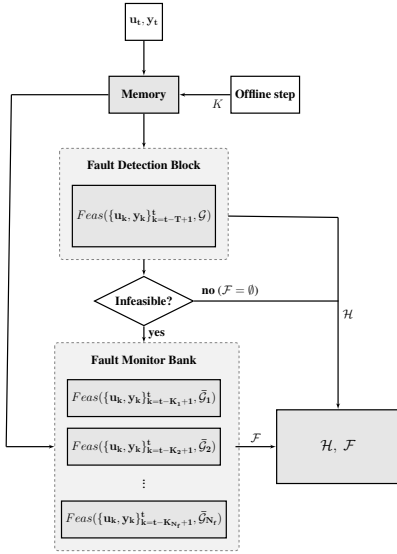


Fig. 1. Block diagram of the proposed FDI scheme.

As we can see, at every time step t , this FDI scheme acts as a function:

$$[\mathcal{H}, \mathcal{F}] = \psi(\{\mathbf{u}_k, \mathbf{y}_k\}_{k=t-K+1}^t, \mathcal{G}, \{\bar{\mathcal{G}}_j\}_{j=1}^{N_f}), \quad (7)$$

where \mathcal{H} is either 0 or 1 to indicate healthy or faulty behaviors, and \mathcal{F} either indicates the fault model that is active, or claims that none of the fault models matches the faulty behavior.

Remark 3: In some practical examples, Assumptions A1 and A2 may not be satisfied, i.e., the FDI approach is not guaranteed to detect and isolate the given fault models. However, the FDI approach can be simply modified such that \mathcal{F} outputs either the set of faults that matches the data (because some fault models may not be isolable) along with their corresponding ‘likelihoods’ in terms of their distinguishability indices, or the empty set if none of the models matches the data.

A. Detection and Isolation Delays

In this section, we describe the notion of delays in detection and isolation of faults, and provide theoretical bounds

on these delays using detectability and isolability indices.

Definition 7: (Detection/Isolation Delay) Detection/isolation delay is the number of time samples it takes from the occurrence of the fault to its detection/isolation. We denote detection and isolation delays with τ_T and τ_I , respectively.

Proposition 4: Given a T_i -detectable pair of system and fault models $(\mathcal{G}, \bar{\mathcal{G}}_i)$, the detection delay of the proposed fault detection scheme is bounded by T_i . In addition, the isolation delay of a pair of $I_{i,j}$ -isolable fault models $(\bar{\mathcal{G}}_i, \bar{\mathcal{G}}_j)$ is bounded by $I_{i,j}$.

Proof: Direct consequence of definitions. ■

Theorem 2: The detection delay for fault $\bar{\mathcal{G}}_i$ using FDI scheme proposed in Section V is bounded by T_i , and the isolation delay is bounded by $K_i = \max\{\bar{I}_i, T_i\}$.

Proof: Assume fault i occurs at time t^* . The FDI approach implements model invalidation with time horizon size of $T \geq T_i$. At the time $t^* + T_i - 1$, the input-output trajectory that is fed to the model invalidation contains a length T_i trajectory that is in $\mathcal{B}^{T_i}(\bar{\mathcal{G}}_i)$. By T_i -detectability of $\bar{\mathcal{G}}_i$, this trajectory cannot be generated by \mathcal{G} . Therefore, the model will be invalidated at most by observing T_i data points from fault i . This concludes the proof for bound on detection delay. For isolation, the FDI approach requires detection first, and in the worst case detection will occur in T_i steps. On the other hand, if we observe any trajectory from t^* to $t^* + \bar{I}_i - 1$ that is generated by fault i , it is not in $\mathcal{B}^{\bar{I}_i}(\bar{\mathcal{G}}_j)$, $j \neq i$. This is because $\bar{I}_i \geq I_{i,j}$, $j \neq i$. Hence, the fault will be isolated in at most \bar{I}_i observations of the fault. Considering that the fault needs to be detected first, the isolation delay is bounded by $K_i = \max\{\bar{I}_i, T_i\}$. This concludes the proof. ■

B. Adaptive Fault Isolation

The bound on isolation delays represents the worst case scenario, where the data created by a fault model falls within the behavior of some other models up until the very last time step. However, this is not the case in most applications, where the faults can be isolated much prior to this bound. Here, in this section, we propose an adaptive fault isolation scheme that reduces isolation delay, which is based on the idea of validation of only one of the fault models. Since the data prior to the time of detection will most probably invalidate all the fault models, we propose to reduce isolation delays by using an adaptive receding horizon that considers only the data starting from the detection time (fixed horizon lower bound) with increasing horizon until only one fault model matches or validates the data. In practice, we can achieve this by considering model invalidation problems for each of the fault models with the adaptive receding horizon until only one fault model remains that matches the data.

Since we assumed that the fault is among the predefined set of models and is persistent, it is guaranteed that the fault will be isolated with this approach. Such an approach has the potential to significantly reduce isolation delays, as we have observed in simulation in Section VI-B (cf. Fig. 4 (bottom)).

VI. ILLUSTRATIVE EXAMPLES

First, we demonstrate in Section VI-A that our new formulations for model invalidation and T -detectability in Prop. 1 and Thm. 1, respectively, are computationally superior to the previous formulation in [19], [28]. Then, we illustrate the performance of the proposed FDI scheme using a numerical model for the Heating, Ventilating, and Air Conditioning (HVAC) system that is proposed in [33] in Section VI-B. Moreover, we provide a numerical example in Section VI-C to illustrate the practical merits of the distinguishability index that was introduced in Section III-B. All the simulations in this section are implemented on a 3.5 GHz machine with 32 GB of memory that runs Ubuntu. For the implementation of the MILP feasibility check problems, we utilized YALMIP [34] and Gurobi [29]. All the approaches and examples are implemented in MATLAB, and are partially available in the MI4Hybrid¹ toolbox.

A. Run-Time Comparison

In this section, we compare the run-time for the formulations proposed in this paper with the one of proposed in [28]. Consider a hidden-mode switched affine model, \mathcal{G} , with admissible sets $\mathcal{X} = \{\mathbf{x} \mid \|\mathbf{x}\| \leq 11\}$, $\mathcal{U} = \{\mathbf{u} \mid \|\mathbf{u}\| \leq 1000\}$ and $\mathcal{E} = \{\boldsymbol{\eta} \mid \|\boldsymbol{\eta}\| \leq 0.1\}$. We assume there is no process noise. We also assume a fixed $B = [1 \ 0 \ 1]^T$ and $C = [1 \ 1 \ 1]$ for all modes. The system matrices of the modes are:

$$\mathbf{A}_1 = \begin{bmatrix} 0.5 & 0.5 & 0.5 \\ 0.1 & -0.2 & 0.5 \\ -0.4 & 0.6 & 0.2 \end{bmatrix}, \mathbf{f}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \mathbf{A}_2 = \begin{bmatrix} 0.5 & 0.5 & 0.5 \\ -0.3 & -0.2 & 0.3 \\ 0.1 & -0.3 & -0.5 \end{bmatrix}, \mathbf{f}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$$

$$\mathbf{A}_3 = \begin{bmatrix} 0.5 & 0.2 & 0.6 \\ 0.2 & -0.2 & 0.2 \\ -0.9 & 0.7 & 0.1 \end{bmatrix}, \mathbf{f}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}.$$

In addition, consider a fault model, \mathcal{G}^f , with the following parameters:

$$\mathbf{A}^f = \begin{bmatrix} 0.8 & 0.7 & 0.6 \\ 0.1 & -0.2 & 0.3 \\ -0.4 & 0.3 & -0.2 \end{bmatrix}, \mathbf{B}^f = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \mathbf{f}^f = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}.$$

The implementation of the T -detectability approach proves this fault to be 12-detectable for the system. We first randomly generate input-output trajectories (5 for each time horizon length) from \mathcal{G}^f . We then apply model invalidation approach implemented using the proposed formulation in Prop. 1 and the one in [19], [28]. The average run-time for each time horizon length as well as the standard deviation of run-times for both formulations are illustrated in Fig. 2. Clearly, the results indicate the superiority of the proposed formulation to the one in [19], [28]. Similar improvements were also observed for the proposed T -detectability formulation in Thm. 1 when compared to [19], [28] (plots are omitted for brevity).

B. Fault Diagnosis in HVAC Systems

In [33], a single-zone HVAC system in cooling mode (cf. schematic in Fig. 3) is considered. This HVAC system is

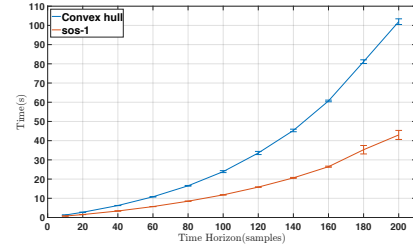


Fig. 2. The average execution time (with standard deviations) for the invalidation of data generated by \mathcal{G}^f on various time horizons.

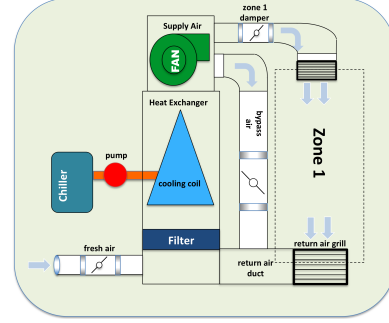


Fig. 3. Schematic of a single-zone HVAC system.

represented by an affine time variant model as follows:

$$\begin{pmatrix} \dot{T}_{TS} \\ \dot{W}_{TS} \\ \dot{T}_{SA} \end{pmatrix} = \begin{pmatrix} -\frac{f}{V_s} & \frac{h_{fg}f}{C_p V_s} & \frac{f}{V_s} \\ 0 & -\frac{f}{V_s} & 0 \\ 0.75 \frac{f}{V_{he}} & -0.75 \frac{h_{fw}}{C_p V_{he}} & -\frac{f}{V_{he}} \end{pmatrix} \begin{pmatrix} T_{TS} \\ W_{TS} \\ T_{SA} \end{pmatrix} + \begin{pmatrix} -\frac{h_{fg}f}{C_p V_s} W_s + \frac{4}{C_p V_s} (Q_o - h_{fg} M_o) \\ \frac{f}{V_s} W_s + \frac{M_o}{\rho V_s} \\ \frac{f}{4V_{he}} (T_o - \frac{h_{fw}}{C_p} W_o) + \frac{f h_{fw}}{C_p V_{he}} W_s - 6000 \frac{gpm}{\rho C_p V_{he}} \end{pmatrix}, \quad (8)$$

where f , gpm , M_o and Q_o are time varying parameters. The parameters of the model are defined in Table I.

In this model, the flow rate of air and chilled water are considered as inputs. We assume that the fan is always turned on and the flow rate is fixed at 17000 ft³/min. Furthermore, it is assumed that the chiller is either turned on with a fixed chilled water flow of 58 gal/min or turned off. In order to linearize the model, an augmented state-space model with additional states Q_o and M_o is obtained in [33], which will be used to convert the non-linear model represented by (8)

TABLE I
PARAMETERS OF THE MODEL

Parameter	Description	Value
h_w	Enthalpy of liquid water	180 (Btu/lb)
h_{fg}	Enthalpy of water vapor	1078.25 (Btu/lb)
W_o	Humidity ratio of outdoor air	0.018 (lb/lb)
W_s	Humidity ratio of supply air	0.007 (lb/lb)
W_{TS}	Humidity ratio of thermal space	state variable
C_p	Specific heat of air	0.24 (Btu/lb.°F)
T_o	Temperature of outdoor air	85 (°F)
T_{SA}	Temperature of supply air	state variable (°F)
T_{TS}	Temperature of thermal space	state variable (°F)
V_s	Volume of thermal space	58464 (ft ³)
V_{he}	Volume of heat exchange space	60.75 (ft ³)
M_o	Moisture load	166.06 (lb/hr)
Q_o	Sensible heat load	289897.5 (Btu/hr)
ρ	Air mass density	0.074 (lb/ft ³)
f	Volumetric flow rate of air	input (ft ³ /min)
gpm	Flow rate of chilled water	input (gal/min)

¹<https://github.com/data-dynamics/MI4Hybrid>

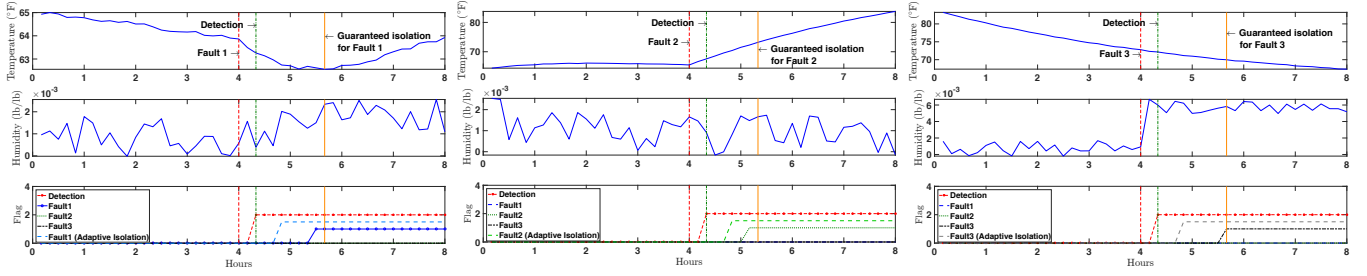


Fig. 4. The outputs of three scenarios (top two rows) and detection, isolation and adaptive isolation signals for all faults (bottom row)

to an SWA model parameterized by the following matrices after discretization with a sampling time of 10 minutes:

$$A_1 = A_2 = \begin{pmatrix} 0.98 & 229.63 & 0.001 & 0 & -0.0035 \\ 0 & 0.94 & 0 & 0 & 0 \\ 0.74 & -360.61 & 0.0008 & 0 & -0.0030 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \quad (9)$$

$$C_1 = C_2 = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \end{pmatrix}, \quad f_1 = \mathbf{0}, \quad f_2 = \begin{pmatrix} 0.3886 \\ 0.0001 \\ -22.576 \\ 0 \\ 0 \end{pmatrix}.$$

In addition, the HVAC model is represented by $\mathcal{G}_H = (\mathcal{X}, \mathcal{E}, \mathcal{U}, \{G_i\}_{i=1}^2)$, where $\mathcal{X} = \{\mathbf{x} \mid [0 \ 0 \ 0 \ 289800 \ 50]^T \leq \mathbf{x} \leq [120 \ 0.02 \ 120 \ 289950 \ 70]^T\}$, $\mathcal{E} = \{\boldsymbol{\eta} \mid |\boldsymbol{\eta}| \leq [0.1 \ 0.001]^T\}$ and $\mathcal{U} = \emptyset$. The last two bounds on the states are for the augmented states, which are assumed to stay within a small range of their equilibria. The first mode corresponds to chiller being ‘on’ and the second mode represents the model when it is ‘off’. The controller keeps the temperature in the comfort zone of 65–75°F by turning the chiller on and off. The control signals are not observed by the FDI scheme.

We consider three fault scenarios:

- 1) Faulty fan: The fan rotates at half of its nominal speed.
- 2) Faulty chiller water pump: The pump is stuck and spins at half of its nominal speed.
- 3) Faulty humidity sensor: The humidity measurements are biased by an amount of +0.005.

The proposed approach for T -detectability and I -isolability gives us the following results:

TABLE II
DETECTABILITY AND ISOLABILITY INDICES

$T_1 = 15$	$T_2 = 5$	$T_3 = 5$	$I_{1,2} = 8$	$I_{1,3} = 10$	$I_{2,3} = 5$
------------	-----------	-----------	---------------	----------------	---------------

In order to illustrate the growth in the distinguishability indices δ^* , we plot its trend in Fig. 5 (left) as the time horizon increases for T -detectability of fault 1 and I -isolability of faults 1 and 2. The plot shows that the distinguishability index we introduced does indeed deliver a nice measure of how far two models are from detectability or isolability, and at the same time, it allows us to estimate the size of time horizon, T or I , to achieve T -detectability or I -Isolability.

Now that we have calculated I and T , the maximum of the two is taken to be the time horizon size to implement the bank of model invalidation monitors, i.e., with $K = 15$. Next, we consider 3 scenarios, where for each scenario i

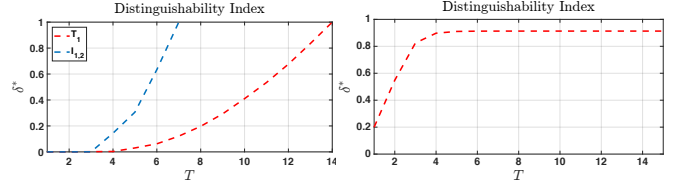


Fig. 5. Distinguishability index as a function of length of time horizon: Left: increase in detectability index for fault 1, T_1 and isolability index of faults 1 and 2, $I_{1,2}$, for HVAC example, Right: Nonlinear increase with a plateau at around $T = 5$, for numerical example described by (10).

($i \in \{1, 2, 3\}$), we generate data from the nominal system for four hours and from fault i afterwards. The times at which the faults occur and their detection times, as well as the upper bounds on isolation delays are indicated in Fig. 4 (top and middle), which show the output trajectories for each scenario. Furthermore, we plot in Fig. 4 (bottom) the detection and isolation signals for all three faults to show that only the occurred fault is isolated in all scenarios before their upper bounds are exceeded, and that the proposed adaptive isolation scheme reduces the isolation delay, as desired.

C. Distinguishability Index

To illustrate the practical use of the distinguishability index, δ^* , we consider two synthetic SWA models \mathcal{G} and $\bar{\mathcal{G}}$ subject to measurement and process noise, given by

$$\mathcal{G} : \begin{cases} A_1 = \begin{pmatrix} 0.15 & 0 & 0.1 \\ 0 & 0.1 & 0.22 \\ 0.2 & 0.1 & 0 \end{pmatrix}, A_2 = \begin{pmatrix} 0 & 0 & 0.12 \\ 0.1 & 0 & 0 \\ 0.1 & 0.15 & 0.1 \end{pmatrix}, \\ C_1 = C_2 = I, f_1 = \begin{pmatrix} 0 \\ 0.7 \\ -0.8 \end{pmatrix}, f_2 = \begin{pmatrix} -0.2 \\ 0.4 \\ -0.5 \end{pmatrix}, \end{cases} \quad (10)$$

$$\bar{\mathcal{G}} : \begin{cases} \bar{A}_1 = \begin{pmatrix} 0.1 & 0 & 0.1 \\ 0 & 0.1 & 0.2 \\ 0.2 & 0.1 & 0 \end{pmatrix}, \bar{A}_2 = \begin{pmatrix} 0 & 0 & 0.1 \\ 0.1 & 0 & 0 \\ 0.1 & 0.1 & 0.1 \end{pmatrix}, \\ \bar{C}_1 = \bar{C}_2 = I, \bar{f}_1 = \begin{pmatrix} 0.3 \\ 0 \\ 0.9 \end{pmatrix}, \bar{f}_2 = \begin{pmatrix} 0.8 \\ 0.2 \\ 0.3 \end{pmatrix}, \end{cases}$$

where the rest of the parameters are zero. The bounds on the process and measurement noise are set to be 0.2 and 0.24, respectively. Fig. 5 (right) depicts the change of the distinguishability index with increasing T . We observe that the distinguishability index increases nonlinearly and reaches a plateau at a value less than one. In this case, the distinguishability index δ^* provides a practical indication that these two models are very unlikely to be isolable for any finite I . Moreover, if the noise levels are constrained to be below the value of the plateau, then we can be sure that these

faults will be isolable. Hence, the distinguishability index can also be exploited to derive the maximum allowed uncertainty for a system such that certain faults are guaranteed to be detectable or isolable. In turn, this suggests possible measures for ensuring fault detection and isolation through the reduction of noise levels, either with a better choice of sensors or with the use of noise isolation platforms.

VII. CONCLUSION

In this paper, we considered the FDI problem for switched affine models. For fault detection, we proposed new model invalidation and T -detectability formulations using SOS-1 constraints, that are demonstrated to be computationally more efficient and do not require complicated change of variables. Further, we introduced a *distinguishability index* as a measure of separation between the system and fault models and showed that this index is also a practical tool for finding the smallest receding time horizon that is needed for fault detection, as well as for recommending system design changes for ensuring fault detection. Next, we considered the fault isolation problem and introduced the concept of I -isolability that is an analogue to T -detectability. Putting the two together, we have a fault detection and isolation scheme for switched affine models, which guarantees the detection and isolation of faults when certain conditions are met. The scheme is built upon an optimization-based method, which formulates the fault-detection and isolation as MILP feasibility check and optimization problems. The detection and isolation monitors can be implemented independently on several processing units, hence it can be efficiently implemented for a large number of faults. Moreover, we introduced adaptive time horizons in order to isolate faults faster. Finally, we illustrated the efficiency of the proposed approaches with several examples, including with an HVAC system model that is equipped with our FDI scheme.

As future work, we are interested to find system theoretic upper bounds on the time horizon T or I such that the incremental search for the smallest T or I can be efficiently terminated with some formal guarantees.

REFERENCES

- [1] R. Beard. *Failure accommodation in linear systems through self-reorganization*. PhD thesis, MIT, 1971.
- [2] R. Nikoukhan. Guaranteed active failure detection and isolation for linear dynamical systems. *Automatica*, 34(11):1345–1358, 1998.
- [3] R. Nikoukhan and S. Campbell. Auxiliary signal design for active failure detection in uncertain linear systems with a priori information. *Automatica*, 42(2):219–228, 2006.
- [4] P. Frank and X. Ding. Survey of robust residual generation and evaluation methods in observer-based fault detection systems. *Journal of process control*, 7(6):403–424, 1997.
- [5] R. Patton and J. Chen. Observer-based fault detection and isolation: robustness and applications. *Control Engineering Practice*, 5(5):671–682, 1997.
- [6] S. Simani, C. Fantuzzi, and R. J. Patton. *Model-based fault diagnosis in dynamic systems using identification techniques*. Springer Science & Business Media, 2003.
- [7] P. Frank. Advances in observer-based fault diagnosis. In *International Conference on Fault Diagnosis: TOOLDIAG*, 1993.
- [8] H. Sneider and P. M. Frank. Observer-based supervision and fault detection in robots using nonlinear and fuzzy logic residual evaluation. *IEEE Transactions on Control Systems Technology*, 4(3):274–282, 1996.
- [9] R. Isermann. Fault diagnosis of machines via parameter estimation and knowledge processing—tutorial paper. *Automatica*, 29(4):815–835, 1993.
- [10] X. Liu, H. Zhang, J. Liu, and J. Yang. Fault detection and diagnosis of permanent-magnet dc motor based on parameter estimation and neural network. *IEEE Transactions on Industrial Electronics*, 47(5):1021–1030, 2000.
- [11] H. Hammouri, M. Kinnaert, and E. H. El Yaagoubi. Observer-based approach to fault detection and isolation for nonlinear systems. *IEEE Transactions on Automatic Control*, 44(10):1879–1884, 1999.
- [12] S. Paoletti, A. Garulli, J. Roll, and A. Vicino. A necessary and sufficient condition for input-output realization of switched affine state space models. In *47th IEEE Conference on Decision and Control*, pages 935–940, Dec 2008.
- [13] A. Abdo, S. X. Ding, J. Saijai, and W. Damlakhi. Fault detection for switched systems based on a deterministic method. In *IEEE Conference on Decision and Control (CDC)*, pages 568–573, 2012.
- [14] W. Pan, Y. Yuan, H. Sandberg, J. Gonçalves, and G. Stan. Online fault diagnosis for nonlinear power systems. *Automatica*, 55:27–36, 2015.
- [15] S. Narasimhan and G. Biswas. Model-based diagnosis of hybrid systems. *IEEE Transactions on Systems, Man, and Cybernetics, Part A*, 37(3):348–361, 2007.
- [16] S. Campbell and R. Nikoukhan. *Auxiliary signal design for failure detection*. Princeton University Press, 2004.
- [17] J. K. Scott, R. Findeisen, R. D Braatz, and D. M. Raimondo. Input design for guaranteed fault diagnosis using zonotopes. *Automatica*, 50(6):1580–1589, 2014.
- [18] P. Rosa, C. Silvestre, J. S. Shamma, and M. Athans. Fault detection and isolation of ltv systems using set-valued observers. In *IEEE Conference on Decision and Control (CDC)*, pages 768–773, 2010.
- [19] F. Harirchi and N. Ozay. Model invalidation for switched affine systems with applications to fault and anomaly detection. *IFAC ADHS Conference*, 48(27):260–266, 2015.
- [20] F. Harirchi, Z. Luo, and N. Ozay. Model (in)validation and fault detection for systems with polynomial state-space models. In *American Control Conference (ACC)*, pages 1017–1023, July 2016.
- [21] J. Anderson and A. Papachristodoulou. On validation and invalidation of biological models. *BMC bioinformatics*, 10(1):1, 2009.
- [22] N. Ozay, M. Sznajder, and C. Lagoa. Convex certificates for model (in)validation of switched affine systems with unknown switches. *IEEE Transactions on Automatic Control*, 59(11):2921–2932, 2014.
- [23] H. Lou and P. Si. The distinguishability of linear control systems. *Nonlinear Analysis: Hybrid Systems*, 3(1):21–38, 2009.
- [24] P. Rosa and C. Silvestre. On the distinguishability of discrete linear time-invariant dynamic systems. In *IEEE CDC-ECC*, pages 3356–3361, 2011.
- [25] M. Babaali and M. Egerstedt. Observability of switched linear systems. In *International Workshop on Hybrid Systems: Computation and Control*, pages 48–63. Springer, 2004.
- [26] E. Beale and J. Forrest. Global optimization using special ordered sets. *Mathematical Programming*, 10(1):52–69, 1976.
- [27] C. Guéret, C. Prins, and M. Sevaux. Applications of optimization with xpress-mp. *contract*, page 00034, 1999.
- [28] F. Harirchi and N. Ozay. Guaranteed model-based fault detection in cyber-physical systems: A model invalidation approach. [arXiv:1609.05921 \[math.OC\]](https://arxiv.org/abs/1609.05921), 2016.
- [29] Inc. Gurobi Optimization. Gurobi optimizer reference manual, 2015.
- [30] IBM ILOG CPLEX. V12. 1: User’s manual for cplex. *International Business Machines Corporation*, 46(53):157, 2009.
- [31] F. Harirchi, S.Z. Yong, E. Jacobsen, and N. Ozay. Active model discrimination with applications to fraud detection in smart buildings. In *IFAC World Congress, Toulouse, France*, 2017.
- [32] E. Jacobsen, F. Harirchi, S.Z. Yong, and N. Ozay. Optimal input design for affine model discrimination with applications in intention-aware vehicles. *arXiv preprint arXiv:1702.01112*, 2017.
- [33] B. Argüello-Serrano and M. Vélez-Reyes. Nonlinear control of a heating, ventilating, and air conditioning system with thermal load estimation. *IEEE Transactions on Control Systems Technology*, 7(1):56–63, 1999.
- [34] J. Löfberg. YALMIP: A toolbox for modeling and optimization in MATLAB. In *CACSD Conference*, Taipei, Taiwan, 2004.